ORIGINAL RESEARCH

Association Mapping for Grain Quality in a Diverse Sorghum Collection

Sivakumar Sukumaran, Wenwen Xiang, Scott R. Bean, Jeffrey F. Pedersen, Stephen Kresovich, Mitchell R. Tuinstra, Tesfaye T. Tesso, Martha T. Hamblin, and Jianming Yu*

Abstract

Knowledge of the genetic bases of grain quality traits will complement plant breeding efforts to improve the end-use value of sorghum [*Sorghum bicolor* (L.) Moench]. Candidate gene association mapping was used on a diverse panel of 300 sorghum accessions to assess marker–trait associations for 10 grain quality traits measured using the single kernel characterization system (SKCS) and near-infrared reflectance spectroscopy (NIRS). The analysis of the accessions through 1290 genomewide single nucleotide polymorphisms (SNPs) separated the panel into five subpopulations that corresponded to three major sorghum races (durra, kafir, and caudatum), one intermediate race (guinea-caudatum), and one working group (zerazera-caudatum). These subpopulations differed in kernel hardness, acid detergent fiber, and total digestible nutrients. After model testing, association analysis between 333 SNPs in candidate genes and/or loci and grain quality traits resulted in eight significant marker–trait associations. A SNP in starch synthase *IIa* (*SSIIa*) gene was associated with kernel hardness (KH) with a likelihood ratio-based R^2 (R_{LR}^2) value of 0.08, a SNP in starch synthase (*SSIIb*) gene was associated with starch content with an R_{LR}^2 value of 0.10, and a SNP in loci *pSB1120* was associated with starch content with an R_{LR}^2 value of 0.09.

Published in The Plant Genome 5:126–135.
doi: 10.3835/plantgenome2012.07.0016
© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

Sorghum is an important cereal crop used as human food in the semiarid tropics of the African and Asian continents by approximately 500 million people (De Wet, 1978). It is a gluten-free cereal used as whole grain as well as ground flour and it is a source of energy, protein, vitamins, minerals, and nutraceuticals such as antioxidant phenolics and cholesterol-lowering waxes (Taylor et al., 2006). Grain quality varies among different types of sorghum and their cultivated environments. Genetic improvement of grain quality can help sorghum to adapt to varying demands for end-use products.

Grain quality is differentiated by biochemical and physical characteristics in sorghum. Kernel hardness (KH) affects grain mold resistance (Jambunathan et al., 1992), grain storage ability, insect resistance (Bueso et al., 2000), milling behavior (Suhendro et al., 2000), flour particle size, cooking properties (Anglani, 1998; Bettge et al., 2000), and parameters such as adhesion, cooked grain texture, alkali gel stiffness (Cagampang and Kirleis, 1984), porridge quality (Akingbala and Rooney, 1987), and production of

S. Sukumaran, W. Xiang, T.T. Tesso, and J. Yu, Dep. of Agronomy, Kansas State Univ., Manhattan, KS 66506; S.R. Bean, USDA-ARS, Grain Storage and Structural Research Unit, Manhattan, KS 66502; J.F. Pedersen, USDA-ARS, Grain, Forage, and Bioenergy Research, Univ. of Nebraska, Lincoln, NE 68583; S. Kresovich, Dep. of Biological Sciences, Univ. of South Carolina, Columbia, SC 29208; M.R. Tuinstra, Dep. of Agronomy, Purdue Univ., West Lafayette, IN 47907; M.T. Hamblin, Institute for Genomic Diversity, Cornell Univ., Ithaca, NY 14853. Received 11 July 2012. *Corresponding author (jyu@ksu.edu).

Abbreviations: ADF, acid detergent fiber; BIC, Bayesian information criterion; CP, crude protein; K, kinship; *k*, subpopulation; KD, kernel diameter; KH, kernel hardness; KW, kernel weight; MAF, minor allele frequency; NIRS, near-infrared reflectance spectroscopy; NJ, neighbor joining; nMDS, nonmetric multidimensional scaling; PC, principal component; PCA, principal component analysis; Q, population structure; Q-Q, quantile–quantile; R_{LR}^2, likelihood ratio-based R^2; SCP, sorghum conversion program; SKCS, single kernel characterization system; SNP, single nucleotide polymorphism; TDN, total digestible nutrients.

high-quality couscous granules (Aboubacar and Hamaker, 1999). Sorghum kernels are round and small in size and vary from about 3 to 4 mm in diameter. Variation in kernel diameter (KD) exists among cultivars (Wills and Ali, 1983a). Large sorghum kernels with corneous endosperm are usually preferred for human consumption and are associated with desirable physical and chemical quality parameters such as high protein concentration, low ash, high milling yields, high water absorbance flour, bright white color, and large particle size (Lee et al., 2002). Small-kernel sorghum that is more likely to be harder and more difficult to mill is not popular in the grain market (Wills and Ali, 1983b). Kernel weight (KW) contributes to grain yield, and its components (kernel moisture content and kernel density) are correlated with milling value. Sorghum grain contains higher levels of acid detergent fiber (ADF) than yellow corn (*Zea mays* L.), and high-tannin varieties contain higher amounts of ADF than non-tannin sorghum (Douglas et al., 1990). Chemical quality parameters such as crude protein (CP), fat, P, starch, and total digestible nutrients (TDN) directly influence sorghum nutritional value. Starch content in sorghum kernels affects the consistency of thick porridge, cooked couscous firmness, and rollability (Beta et al., 2001).

Genetic mapping of grain quality traits has been conducted in different cereal crops such as maize (Cook et al., 2012; Wilson et al., 2004), rice (*Oryza sativa* L.) (Tian et al., 2009), wheat (*Triticum aestivum* L.) (Bordes et al., 2011; Reif et al., 2011), and sorghum (de Alencar Figueiredo et al., 2010). Starch is one of the most important grain quality parameters in cereals that provide the basis of subsistence for world population. Four enzymes, adenosine diphosphate glucose pyrophosphorylase, starch synthase, starch branching enzyme, and starch debranching enzyme, catalyze starch biosynthesis in cereals (Preiss, and Sivak, 1998). Given that the pathways and enzymes related to grain quality are similar in cereals, it was not unexpected to have similar results through earlier mutational studies and recent association studies at the population level.

A community resource sorghum diversity panel was recently established by a collection of sorghum accessions representing all major cultivated races, including lines from the Sorghum Conversion Program (SCP), elite breeding lines, and their progenitors from all around the United States (Casa et al., 2008). These accessions are available on request from the National Plant Germplasm System (http://www.ars-grin.gov/npgs/index.html) at Plant Genetic Resources Conservation Unit, Griffin, GA. The SCP converted tropical lines to photoperiod-insensitive short plants (Stephens et al., 1967). The level of population structure and familial relatedness in this diversity panel was previously assessed using 47 simple sequence repeat markers (Casa et al., 2008). Another study analyzed 216 SCP lines using 434 single nucleotide polymorphisms (SNPs) and classified the lines into four subpopulations that corresponded closely to four major races of sorghum. A combined analysis of the breeding lines and lines from the SCP program has not been conducted with a large number of markers; furthermore, genetic mapping studies will complement breeders' efforts to improve grain quality in sorghum. The present research was undertaken to identify marker–trait associations for grain quality traits in sorghum.

MATERIALS AND METHODS

Plant Germplasm and Phenotypic Characterization

Three hundred lines from the sorghum diversity panel, including 251 SCP lines, and 49 important breeding lines and their progenitors from the United States served as the genetic material for this study. The sorghum accessions were planted with randomized complete block design in Manhattan, KS, and West Lafayette, IN, with two replications in 2007 and 2008. Seeds harvested from 10 selfed sorghum heads were analyzed for grain quality using the single kernel characterization system (SKCS) (Martin et al., 1993) and near-infrared reflectance spectroscopy (NIRS) (Pasquini, 2003). The SKCS provided data on KH, KD, and KW and NIRS provided data on ADF, Ca, CP, fat, P, starch, and TDN.

Single Kernel Characterization System

The SKCS was the device used to measure physical properties of sorghum kernels such as KH and size characteristics (Bean et al., 2006; Pedersen et al., 1996). Seeds of 287 lines from 2 yr and two replications were analyzed through SKCS. Three hundred individual grains were crushed between a serrated rotor and a crescent, and parameters for KH, KD, and KW were estimated and reported. Kernel hardness was reported as kernel hardness index.

Near-Infrared Reflectance Spectroscopy

Near-infrared reflectance spectroscopy utilizes the near-infrared region of the electromagnetic spectrum (about 800–2500 nm) to determine the concentration of physical and chemical constituents in agricultural materials (Pasquini, 2003). Near-infrared reflectance spectroscopy was used to predict the amount of ADF, CP, fat, Ca, P, starch, and TDN in sorghum kernels. A total of 15 g of seed were ground in a UDY cyclone mill (UDY Corporation) with a 1-mm screen, a stainless steel grinding ring, and an aluminum impeller. Two hundred sixty-nine lines from two replications in Manhattan (2007) were scanned using a Foss NIRSystem 6500 monochromator (NIR Systems Inc.). High R^2 values were obtained for various traits using a validation set of 52 samples. The R^2 for starch, CP, fat, ADF, and P contents were 0.99, 0.98, 0.91, 0.88, and 0.88, respectively. On the basis of the statistical parameters mentioned above, NIRS was demonstrated to be efficient and accurate in predicting chemical grain quality traits in this panel.

Genotyping and Candidate Genes

Two different genotyping assays were conducted: (i) a genomewide assay of 1536 SNPs (Yu et al., 2011) and (ii) a candidate gene and/or loci assay of 384 SNPs (Murray

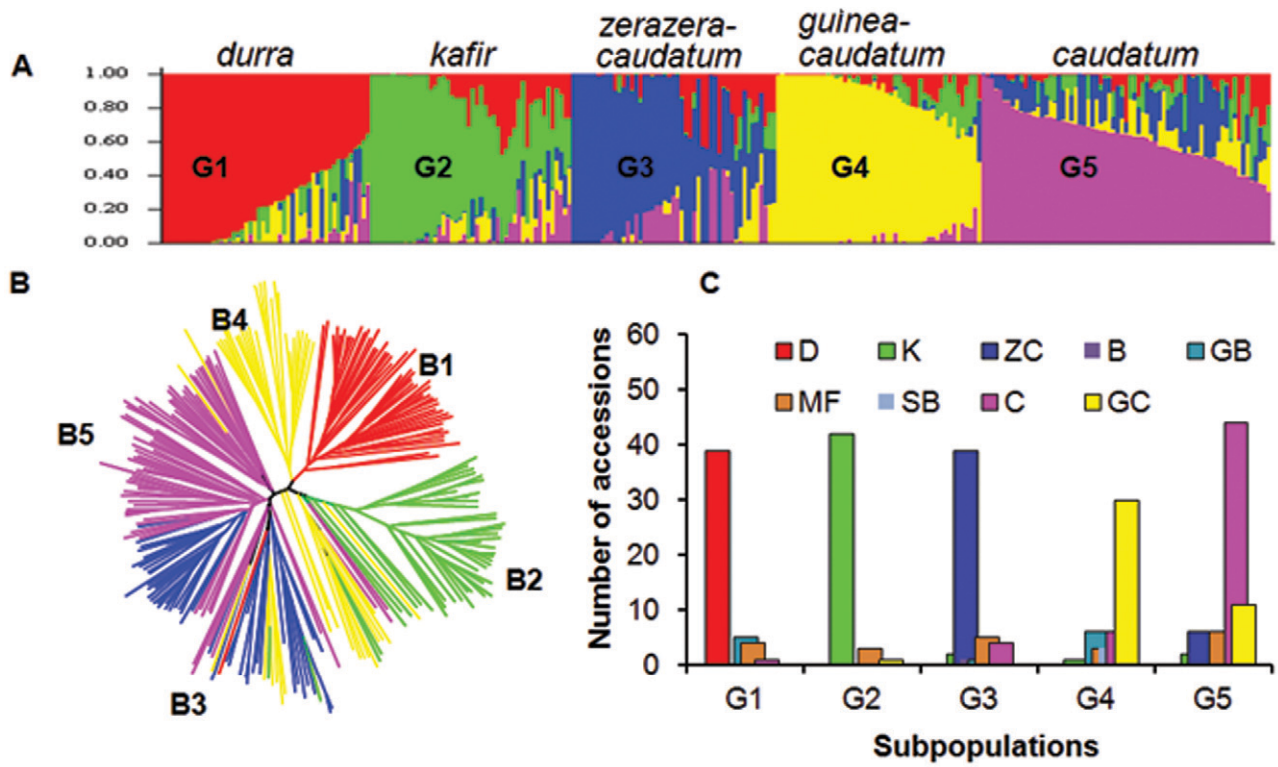

Figure 1. Diversity analysis of the sorghum accessions. (A) STRUCTURE (Pritchard et al., 2000) results: five subpopulations (G) corresponded to races. (B) Neighbor joining tree: branches (B1–B5) generally agreed to subpopulations (G) based on STRUCTURE results. (C) Number of accessions from specific races within each subpopulation. D, durra; K, kafir; ZC, zerazera-caudatum; B, bicolor; GB, guinea-bicolor; MF, milo-feterita; SB, sudanese-broomcorn; C, caudatum; GC, guinea-caudatum.

et al., 2009). The 1536 SNP assay was designed to achieve maximum genome coverage. The average distance between SNPs was 400 kb except in the centromere regions. The 384 SNP assay was developed from SNPs discovered in previously published studies (Hamblin et al., 2004, 2005, 2006, 2007), starch pathways (Hamblin et al., 2007), sucrose pathways (Murray et al., 2009), and carotenoid pathways (Salas Fernandez et al., 2009). Out of 226 loci represented in the 384 SNP assay, 39 loci were candidate genes from starch, sucrose, and carotenoid pathways and the remainder were candidate loci distributed across 10 chromosomes. An Illumina GoldenGate assay was used to genotype the samples. Out of the 1536 SNP assay, 1290 SNPs and, out of the 384 SNP assay, 333 SNPs were successful and polymorphic. The program fastPHASE was used to impute missing data (Scheet and Stephens, 2006).

Statistical Analysis

DNA Marker Profile

PowerMarker version 3.25 (Liu and Muse, 2005) was used to calculate Chord distance (Cavalli-Sforza and Edwards, 1967) among accessions. It was also used to compute molecular diversity statistics and to construct the neighbor joining (NJ) tree with 100 replications of bootstrapping.

Population Structure Analysis

The program STRUCTURE, version 2.2.3 (Pritchard et al., 2000), was used to detect population structure and assign individuals to subpopulations. The STRUCTURE program was run 10 times for each subpopulation (k) value, ranging from 1 to 15, using the admixture model with 20,000 replicates for burn-in and 20,000 replicates during analysis. The final subpopulations were determined on the basis of (i) likelihood plot of models, (ii) stability of grouping patterns across 10 runs, (iii) germplasm information or "breeder's knowledge," (iv) cluster analysis (NJ tree), and (v) principal component analysis (PCA). On the basis of this information, we chose k = 5 as the optimal grouping. Out of the 10 runs for k = 5, the run with the highest likelihood value was selected to assign the posterior membership coefficients (population structure [Q]) to each accession (Supplemental Table S1). A graphical bar plot was then generated with the posterior membership coefficients (Fig. 1A), and plots were also plotted for k = 2, 3, 4, and 5 for result interpretation.

To validate the genetic structure and to test marker–trait associations, PCA and nonmetric multidimensional scaling (nMDS) were conducted and kinship (K) matrix was calculated. Principal component analysis was conducted to construct a plot of the most significant axes for grouping pattern variation and to obtain axes for further model testing and association mapping (Patterson et al., 2006; Price et al., 2006; Zhu and Yu, 2009). The combined display of the color-coded subpopulation memberships from STRUCTURE (Pritchard et al., 2000) with other analyses, NJ tree

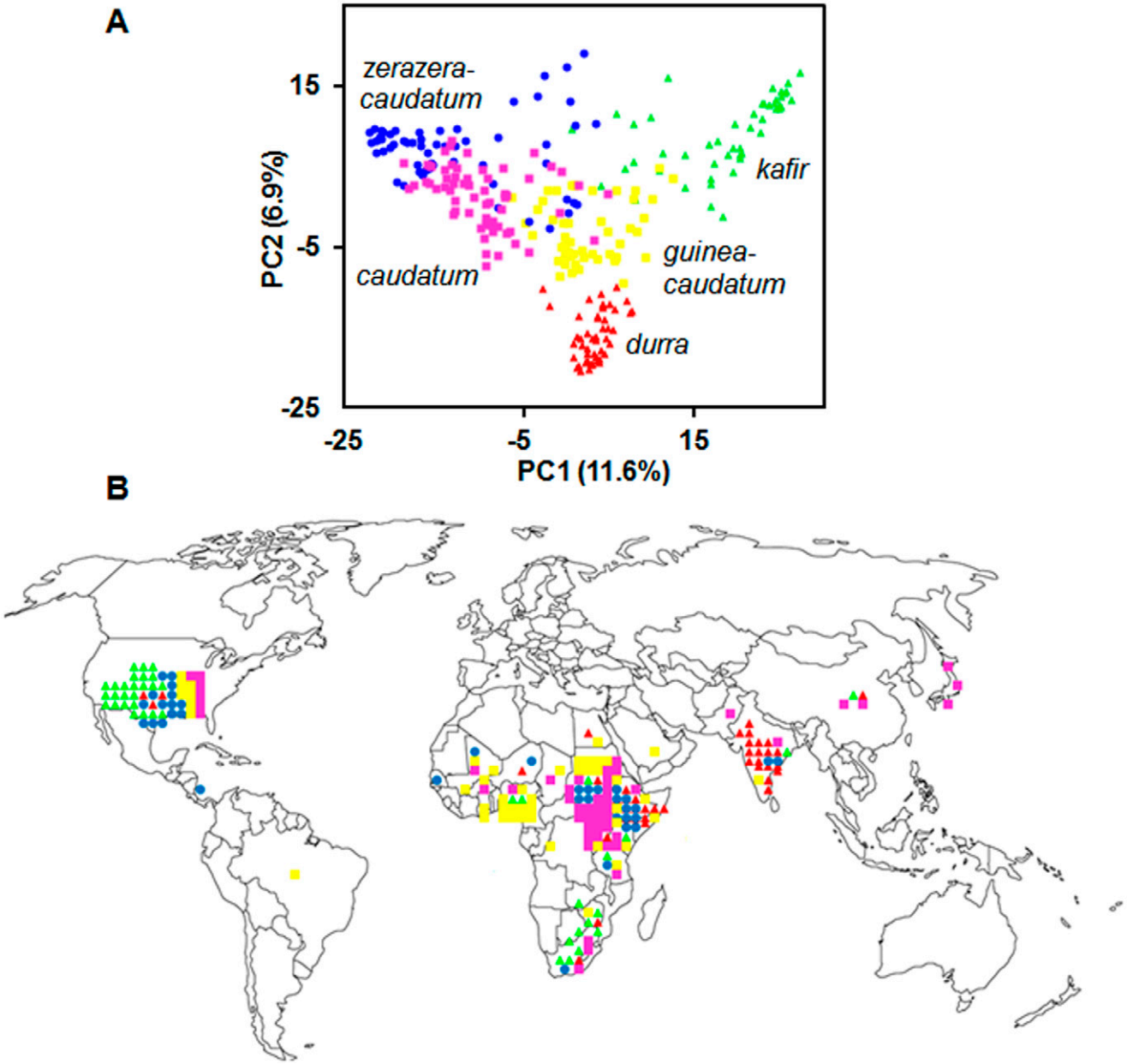

Figure 2. General congruence among principal component analysis (PCA), STRUCTURE (Pritchard et al., 2000) classification, and race classification. (A) Principal component analysis and STRUCTURE classification were consistent. Each color represents a subpopulation based on STRUCTURE results. (B) Geographical origin differences between the subpopulations in sorghum diversity panel shown on world map. Red triangles represent G1, green triangles represent G2, blue circles represent G3, yellow boxes represent G4, and pink boxes represent G5. PC, principal component.

(Fig. 1B) and PCA (Fig. 2A), are shown. Kinship was calculated with SPAGeDi 1.3 (Loiselle et al., 1995; Hardy and Vekemans, 2002).

Model Comparison and Association Analysis

We compared different models to assess the effect of population structure on association mapping of various grain quality traits measured in this diversity panel. Following the previously recommended procedures (Yu et al., 2006; Zhu and Yu, 2009), we tested various mixed models with subpopulation membership percentage (Q), nMDS, and PCA as fixed covariates and kinship as random effect. The dimension of PCA and nMDS were determined for each trait individually. Among all possible models (the simple model, Q, K, PCA, nMDS [Zhu and Yu, 2009], Q plus K [Yu et al., 2006], PCA plus K, and nMDS plus K), the best fit model was determined for each trait based on the Bayesian information criterion (BIC). The selected models were then used to test marker–trait associations between 333 SNPs and 10 grain quality traits. Marker–trait associations were tested in TASSEL (Bradbury et al., 2007) and were also verified in SAS 9.1 software (SAS Institute, 2000). Subsequently, quantile–quantile (Q-Q) plots of the *F*-test statistics for the SNP markers were plotted to assess the

Table 1. Mean, standard deviation, and correlation of grain quality traits across sorghum accessions. The number of accessions used was 247 for kernel hardness (KH), kernel weight (KW), and kernel diameter (KD) and 274 for acid detergent fiber (ADF), Ca, crude protein (CP), fat, P, starch, and total digestible nutrients (TDN).

			Correlation (r)									
Traits	Mean	SD	KH	KW	KD	ADF	Ca	CP	Fat	P	Starch	TDN
KH	78.15	18.97	–									
KW	24.40	5.20	−0.32***	–								
KD	1.70	0.38	−0.33***	0.91***	–							
ADF	4.81	0.91	−0.18**	−0.07	−0.08	–						
Ca	0.06	0.01	0.23***	−0.07	−0.12	0.21***	–					
CP	13.85	1.67	0.18**	0.13	0.05	0.43***	0.31***	–				
Fat	3.25	0.41	0.33***	−0.22***	−0.24***	0.05	0.52***	−0.05	–			
P	0.45	0.05	0.23***	0.04	−0.04	0.28***	0.33***	0.75***	0.20***	–		
Starch	69.26	2.44	−0.09	0.02	0.10	−0.68***	−0.31***	−0.75***	−0.25***	−0.68***	–	
TDN	84.84	1.40	−0.16**	0.07	0.08	−0.99***	−0.21***	−0.43***	−0.05	−0.28***	0.67***	–

**Significant at the 0.01 probability level.

***Significant at the 0.001 probability level.

adequacy of the best model in controlling type I errors (Supplemental Fig. S1). Single nucleotide polymorphisms that passed the threshold of p-value < 10^{-03} were deemed significant if minor allele frequency (MAF) was greater than 5%. The threshold p-value was determined jointly by considering that these are SNPs from candidate genes and their numbers are not very large and by considering the pattern of the Q-Q plot of the selected model and the point at which the observed F-test statistics deviated from the expected F-test statistics. In addition, likelihood-ratio-based R^2 (R_{LR}^2) was calculated for significant SNPs to provide a general measure for the effect of SNPs in mixed-model association mapping of the traits (Sun et al., 2010). Likelihood ratio-based R^2 is a generalized form of R^2 in linear regression model that allows comparisons across models with different random and fixed components.

RESULTS

Population Structure and Genetic Diversity

From the SNP data, the STRUCTURE (Pritchard et al., 2000) analysis revealed five subpopulations (G1, G2, G3, G4, and G5) that contained 49, 46, 52, 49, and 69 accessions, respectively (Fig. 1A). The NJ tree analysis also clustered the data into five branches (Fig. 1B). The color-coded branches support the five subpopulation classification. Each subpopulation closely corresponded to durra, kafir, zerazera-caudatum, guinea-caudatum, and caudatum (Fig. 1C). Subpopulation G1 mainly consists of accessions from the race durra (79.6%), G2 comprises kafir (91.3%), G3 consists of the zerazera-caudatum working group (75%), G4 comprises the guinea-caudatum intermediate race (61.2%), and G5 consists of the caudatum race (63.8%). The genetic group guinea-caudatum is the race guinea in traditional classification. We used information from two earlier studies about the sorghum diversity panel to classify the accessions into genetic groups and races (Brown et al., 2011; Casa et al., 2008).

The results from PCA showed that principal component (PC) 1 explains 11.6% variation in the data by separating G1 from G2, G3, G4, and G5 and PC2 explains 6.9% variation in the data by separating G4 from G1, G2, G3, and G5 (Fig. 2A). The PCA was color-coded based on the structure results and it generally agrees with STRUCTURE (Pritchard et al., 2000) classification of five subpopulations. Even though bicolor is a major race of sorghum, it did not form a specific subpopulation in this diversity panel. We also generated a world map of the accessions based on their sources and/or origins (Fig. 2B). Results from STRUCTURE analysis, NJ tree, and PCA were consistent. Taken together, the sorghum diversity panel was classified into five subpopulations: three main sorghum races (durra, kafir, and caudatum), one intermediate race (guinea-caudatum), and a working group (zerazera-caudatum).

Trait Variation

Data analysis showed a high amount of diversity for grain quality traits. Kernel hardness, KW, and KD from SKCS showed high consistency across years and environments that were recorded from two locations for 2 yr. The repeatability of KH, KW, and KD were 0.79, 0.84, and 0.78, respectively. The correlation coefficients (r) were calculated for all traits. Kernel hardness was significantly correlated with all traits except starch. Kernel weight and KD were positively correlated (r = 0.91). Protein content and P content were positively correlated (r = 0.75). Kernel diameter and fat were negatively correlated (r = −0.24). Crude protein was significantly associated with all traits except KW, KD, and fat. Starch content was negatively correlated with ADF (r = −0.68), Ca (r = −0.31), CP, (r = −0.75), fat (r = −0.25), and P (r = −0.68) (Table 1).

In addition, KH, ADF, and TDN showed significant differences among the five subpopulations (Fig. 3). Caudatum in G5 had the lowest KH (Fig. 3A) and TDN (Fig. 3C) and the highest ADF (Fig. 3B) values. The

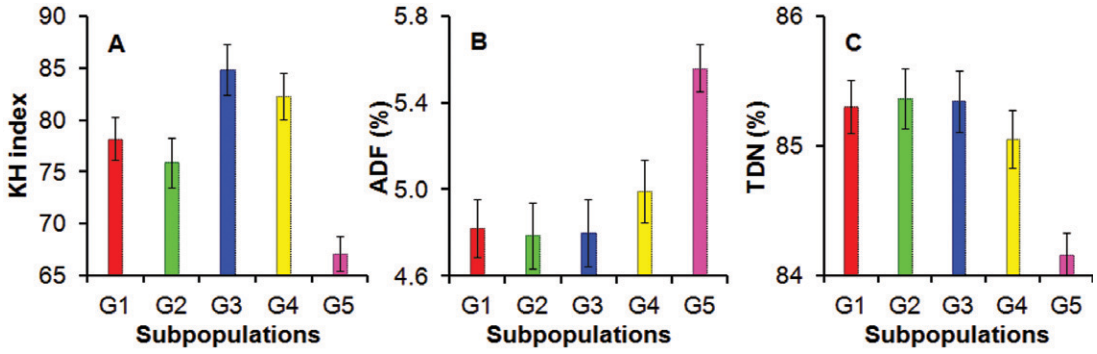

Figure 3. Variations in three grain quality traits among different subpopulations of sorghum diversity panel. (a) Kernel hardness (KH). (b) Acid detergent fiber (ADF). (c) Total digestible nutrients (TDN). The error bar represents the standard error. G1, G2, G3, G4, and G5 consist of 49, 46, 52, 49, and 69 accessions, respectively.

Table 2. Significant single nucleotide polymorphisms (SNPs) in candidate genes associated with grain quality traits.

SNP	Locus and/or gene	Chromosome	Position† (bp)	MAF‡	Trait§	*p*-value (best model)	R_{LR}^2¶ (SNP)	Predicted gene function
SB00214.1	*pSB1700*	3	60,981,646	0.17	KH	1.84×10^{-04}	0.10	Hypothetical protein
SB00214.2	*pSB1700*	3	69,011,869	0.17	KH	1.84×10^{-04}	0.10	Hypothetical protein
SB00116.3	*SSIIa*	10	8,211,818	0.28	KH	7.94×10^{-04}	0.08	Starch synthase IIa
SB00156.1	*pSB0289*	3	58,654,927	0.22	Ca	5.36×10^{-04}	0.06	Serine/threonine-protein kinase
SB00054.1	*PRC1044*	6	59,450,034	0.36	Ca	9.04×10^{-04}	0.06	Hypothetical protein
SB00068.1	*pSB0140*	6	52,406,868	0.07	P	5.83×10^{-04}	0.05	Peptide transporter PTR2
SB00115.3	*SSIIb*	4	58,000,108	0.17	Starch	3.67×10^{-04}	0.10	Starch synthase IIb
SB00086.1	*pSB1120*	3	45,665,466	0.31	Starch	6.19×10^{-04}	0.09	3-ketoacyl-CoA synthase

†Physical location of the SNP in the chromosome in sorghum genome browser v2.39 on Phytozome (http://www.phytozome.net/sorghum [accessed 4 Sept. 2012]; Goodstein et al., 2012).

‡MAF, minor allele frequency.

§KH, kernel hardness.

¶R_{LR}^2, likelihood ratio-based R^2.

accessions that formed G3, zerazera-caudatum, had the highest KH and TDN values followed by guinea-caudatum (G4). Durra (G1) and kafir (G2) accessions had higher KH values than the caudatum (G5) but lower values than zerazera-caudatum (G3) and guinea-caudatum (G4). Caudatum in G5 were significantly different from other subpopulations for these three traits. Other phenotypic traits were not significantly different among the subpopulations.

Marker–Trait Association Analysis

Model comparisons revealed that the mixed model with K matrix was the best model for eight phenotypic traits: KH, KW, KD, ADF, CP, fat, starch, and TDN. The simple model was the best model for testing Ca and the nMDS2 model was the best for testing P. The intersection of phenotypic data (300 accessions) and genotypic data (265 accessions) yielded a combined data set of 200 accessions with both genotypic and phenotypic data. Eight significant marker–trait associations between the SNPs on the candidate genes and grain quality traits were detected after filtering for MAF of 0.05 (Table 2). The Q-Q plots for each phenotype showed that the model tested were effective in controlling type I error (Supplemental Fig. S1). Single nucleotide polymorphisms associated with grain quality traits were checked for the distribution of alleles among subpopulations (Supplemental Fig. S2).

Data analysis revealed three significant SNPs associated with KH, *SB00214.1*, *SB00214.2*, and *SB00116.3*, have *p*-values of 1.84×10^{-04}, 1.84×10^{-04}, and 7.94×10^{-04}, respectively. The consistency of association between significant SNPs and alleles with the trait was checked by plotting the number of alleles in the accessions among five subpopulations (Fig. 4). Accessions with allele *A* in the SNP *SB00214.1* and the accessions with allele *T* in the SNP *SB00214.2* had significantly higher KH values in all subpopulations except in G1 (Fig. 4). *SB00214.1* and *SB00214.2* were located in the locus *pSB1700*, and *SB00116.3* was in starch synthase IIa (*SSIIa*) gene. The values of R_{LR}^2 for these SNPs were 8 to 10%. Except in G2 and G4, accessions with allele *A* in the SNP *SB00116.3* had higher KH values.

Calcium, P, and starch had significant SNPs associated with it (Table 2). Two significant SNPs were associated with Ca. *SB00156.1* and *SB00054.1* were significantly associated with Ca; with *p*-values 5.36×10^{-04} and 9.04×10^{-04}, respectively, each explained 6% of the variation. Except in G4, accessions with allele *A* in the SNP *SB00156.1* had higher Ca content values. In the SNP *SB00054.1*, accessions with allele *C* had higher Ca content values except in G2. *SB00068.1* in candidate loci *pSB0140* in chromosome 6 was significantly associated with P content with a *p*-value of 5.83×10^{-04}. This SNP explained 5% of the variation in P. Accessions with allele

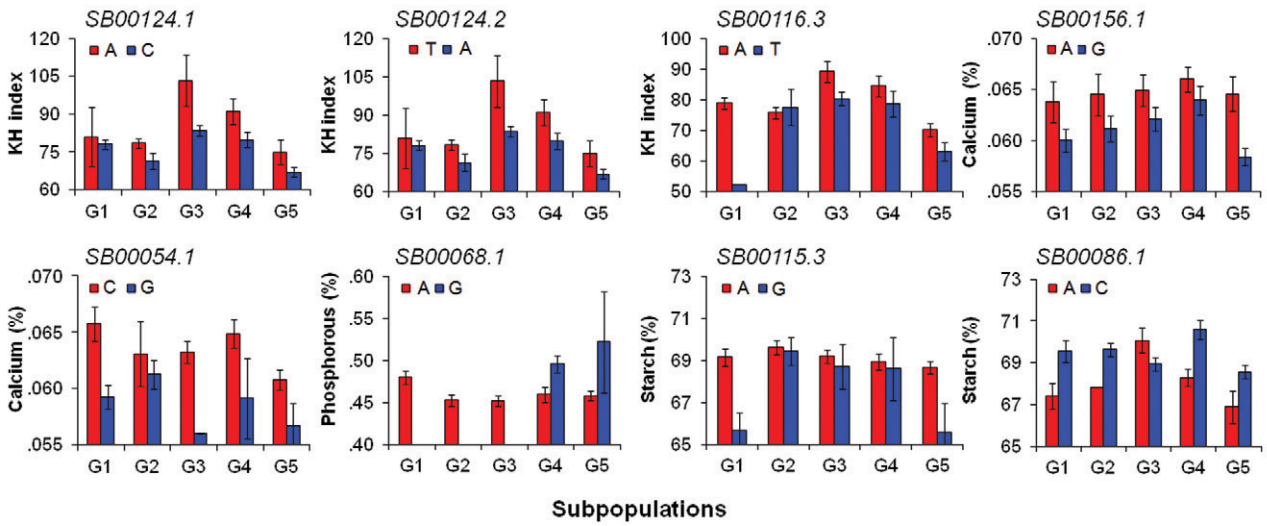

Figure 4. Consistency of single nucleotide polymorphism (SNP) alleles is shown across five subpopulations. Each bar represents mean value of accessions with significant SNP allele. Error bar represents the standard error. KH, kernel hardness.

G had higher P content values in all subpopulations, and the allele *A* was fixed in G1, G2, and G3. Two significant SNPs were associated with starch. *SB00115.3* in candidate gene *SSIIb* was associated with a *p*-value of 3.67×10^{-04}, and *SB00086.1* in *pSB1120* was associated with a *p*-value of 6.19×10^{-04}. The SNP in *SSIIb* explained 10% of the variation and the SNP *SB00086.1* explained 9% of the variation in starch. Accessions with allele *A* in the SNP *SB00115.3* had higher starch content values in G1 and G5. The accession with allele *C* in the SNP *SB00086.1* had higher starch content values except in G3 (Fig. 4).

DISCUSSION

Diversity and Classification in Sorghum

Sorghum is considered to have been domesticated around 5000 to 7000 yr ago in the northeastern part of Africa, the present-day Ethiopia (Jennings and Cock, 1977). Earlier efforts to classify sorghum were mainly based on color of grains and glumes, presence or absence of awn, and stem characteristics. The most complete classification of sorghum was in the early part of the last century (Snowden, 1936). In 1972, another classification based on the spikelet characteristics was proposed (Harlan and De Wet, 1972), and cultivated sorghum was mainly classified into five major races. According to that system, bicolor and guinea races have open panicles, kafir and durra races have compact heads, and caudatum spikelets vary in their head type. Broomcorn generally falls in bicolor type, and feterita is considered to be of caudatum type (Harlan and De Wet, 1972). Notably, no barrier between these sorghum races prevents them from crossing and mixing, so a considerable amount of variation within the five races results from admixture that separates them into about 15 mixed races and nearly 70 working groups (Murthy et al., 1967). Sorghum races also vary in their geographical origin and adaptation (Fig. 2B). The race bicolor is grown almost everywhere in Africa and does not have a characteristic geographical distribution or ecological adaptation. Guineas have hard seeds and are resistant to insects and mold damage under wet conditions. They are grown in the high rainfall areas of West Africa.

We found that the genetic group guinea-caudatum forms a subgroup with high KH values compared with caudatum (Fig. 3A). This genetic group closely corresponds to the race guinea in traditional classification (Brown et al., 2011). The caudatum race is one of the most important; almost all modern hybrid sorghums in the United States are caudatum or are mixed with caudatum. Caudatum has higher yielding ability, bright seed color, and good seed quality. This race is found mixed with other races of sorghum and the working group zerazera-caudatum had the highest KH index (Fig. 3A). The race kafir is found in the southern part of Africa and India. Durra is a drought-tolerant race and is present in India and northern parts of Africa; it is also found mixed with guinea and caudatum.

Earlier efforts to support phenotype-based racial classification in sorghum with genotypic data were successful (Aldrich and Doebley, 1992; Perumal et al., 2007; Folkertsma et al., 2005; Casa et al., 2008; Brown et al., 2011), but there is some disagreement between phenotype- and genotype-based racial classifications. One hypothesis is that, in phenotype-based racial classification of sorghum, the traits used (panicle and spikelet characters) are controlled by a limited number of genomic regions but, in genotype-based classification, a large number of markers are used to classify sorghum races, that is, STRUCTURE (Pritchard et al., 2000) classification is based on random markers distributed across the genome that can capture the genomic variation among sorghum races. However, in an earlier study, 216 SCP lines were classified into four genetic groups that closely corresponded to major traditional races (Brown et al., 2011) except bicolor, and bicolor didn't form a separate subpopulation. Our study showed similar patterns; the

race bicolor was present in G2 through G5 but was not present in the subpopulation G1, which had mostly kafir. Although bicolor is considered the progenitor of all sorghum races (De Wet, 1978), parallel domestication and theories of multiple origin of the sorghum races remain valid. The race kafir might have originated from an early bicolor or a wild race *Sorghum bicolor* (L.) Moench subsp. *verticilliflorum* (Steud.) de Wet ex Wiersema & J. Dahlb. [syn. *Sorghum verticilliflorum* (Steud.) Stapf] (Smith and Frederiksen, 2000). Electrophoresis data from an earlier study suggest that the race kafir was different from the other four major races in protein patterns (Shechter and De Wet, 1975).

A recent study of domestication of the shattering gene in cereals reported multiple *Shattering1* (*Sh1*) alleles for domesticated races in sorghum, and the *Sh1* allele in Tx623 (an important breeding line) found in kafir and bicolor from south and east Africa is different from the alleles found in guinea and durra, which is different from caudatum (Lin et al., 2012). The race bicolor had multiple alleles of the *sh1* gene and didn't have a dominant *sh1* haplotype that indicates wide distribution of this race. The four major races of sorghum probably have multiple independent domestication events (Lin et al., 2012). However, in this sorghum diversity panel zerazera-caudatum formed a separate subpopulation. Our results indicated that this diverse collection of sorghum clustered into five different subpopulations that closely corresponded to three major traditional races, one mixed race, and a working group.

Association Analysis

We followed the unified mixed model approach to account for spurious associations that result from population structure and familial relatedness (Yu et al., 2006). In deciding the best model to test marker–trait associations, we compared and tested different models for the best fit to the phenotypic data. Testing a mixed model with the K matrix in SAS (SAS Institute, 2000) is not a straightforward approach and may encounter convergence problems. The best-fit model (the lowest BIC model) was selected for testing markers for each trait; if each phenotype were not tested with the best model, directly fitting both Q and K may overcorrect population structure and familial relatedness for some traits and result in type II error (Zhu and Yu, 2009).

The SNPs that were significant with MAF < 0.05 were not reported and improved methods are needed to address and identify the true positives (Zhu et al., 2011). The percentage of variation explained was calculated as R_{LR}^2 that is appropriate for mixed model-based association mapping. After controlling for population structure and admixture, we found eight SNPs on the candidate genes to be significantly associated with the grain quality traits. In general, candidate gene association mapping approach complements genomewide association studies and traditional linkage mapping. By using sufficient number of SNPs coupled with careful selection of candidate genes, this approach can establish the gene–trait relationship at the population level.

Marker–Trait Associations

A SNP on the candidate gene *SSIIa* located on chromosome 10 was associated with KH and explained 8% of the variation in the trait. Earlier studies on kernel hardness in sorghum, wheat, and rice suggest that starch content and the distribution of proteins and lipids on the surface of starch granules are important factors in determining grain hardness (Cagampang and Kirleis, 1984; Chen et al., 2012; Guzman et al., 2012; Morris et al., 1994; Yan et al., 2009). Sorghum grain hardness is related to the vitreousness of the grain and the vitreousness is related to amylose content, which is a major component of starch (Cagampang and Kirleis, 1984). The maize homolog of *SSIIa* gene is *sugary2* (*su2*) gene that is in the starch synthesis pathway (Hamblin et al., 2007). Also, the gelatinization temperature in rice is genetically controlled by the *SSIIa* gene that is related to KH (Yan et al., 2010). So these evidences suggest that the gene *SSIIa* from the starch synthesis pathway plays an important role in KH.

Single nucleotide polymorphisms associated with KH (*SB00214.1* and *SB00214.2*) explained 10% of the variation in the trait and they were located in the locus *pSB1700* on chromosome 3. A bioinformatics analysis of the locus revealed that *pSB1700* locus is similar to *sad1* protein in rice and SUN4 domain protein in maize. The translated nucleotide of *pSB1700* had 49% identity with the SUN4 protein in maize. These proteins are present in the inner nuclear membrane in a cell and form a link between other proteins, nucleoskeleton, and cytoskeleton that is important in the structure and shape of a cell (Murphy et al., 2010). Single nucleotide polymorphisms *SB00156.1* and *SB00054.1* associated with Ca content were located in chromosome 3 at 50 and 59 cM, respectively. *SB00156.1* was located in a locus, *pSB0289*, that was predicted to produce serine/threonine-protein kinase. The function of SNP *SB00054.1*, in locus *PRC1044* in chromosome 6, is not known from National Center for Biotechnology Information searches. A SNP, *SB00068.1* in *pSB0140* locus in chromosome 6, explained 5% of the variation in P content. A SNP on the candidate locus *pSB1120* on chromosome 3 was significantly associated with starch content and explained 9% of the variation in the trait. BLAST searches (Altschul et al., 1990) provided the function of this locus as a gene producing 3-ketoacyl-CoA synthase. These five SNP–trait associations are novel associations.

A SNP on the starch synthase IIb (*SSIIb*) gene on chromosome 4 was found to be significantly associated with starch content and explained 10% of the variation in the trait. The maize homolog of this gene is *SSIIb*. Starch synthase is an enzyme required for starch synthesis in the endosperm of cereals (Fujita et al., 2011). Candidate gene association mapping in maize (Wilson et al., 2004) and rice (Tian et al., 2009) suggest that the starch synthase is an important enzyme in determining starch content and quality in cereals. The genes from starch synthesis pathways form a regulatory network and

influence grain quality parameters (Tian et al., 2009). In sweet wheat, in the absence of the granule bound starch synthase II, starch is not formed in its kernels (Shimbata et al., 2011). The findings about the trait differences among different subpopulations and the identified SNPs from the present study can be further exploited in improving grain quality in sorghum and related cereals.

Supplemental Information Available

Supplemental material is available at http://www.crops.org/publications/tpg.

Supplemental Fig. S1. Quantile–quantile plots of the 10 grain quality traits with 1523 single nucleotide polymorphism (SNP) markers. The quantile–quantile (Q-Q) plots showed the control of type I error by the selected models. KH, kernel hardness; KW, kernel weight; KD, kernel diameter; ADF, acid detergent fiber; CP, crude protein; TDN, total digestible nutrients.

Supplemental Fig. S2. Variation in the frequency of significant single nucleotide polymorphisms (SNPs) associated with grain quality traits across the five subpopulations. Each bar represents the number of alleles for each SNP in the subpopulation.

Acknowledgments

This work was supported by the Agriculture and Food Research Initiative Competitive Grant (2011-03587) from the USDA National Institute of Food and Agriculture, the USDA-ARS, the Kansas Grain Sorghum Commission, and the Kansas State University Center for Sorghum Improvement. Names are necessary to report factually on available data; however, USDA neither guarantees nor warrants the standard of the product, and the use of the name by the USDA implies no approval of the product to the exclusion of others that may also be used. This is contribution number 12-463-J from the Kansas Agricultural Experiment Station, Manhattan, KS.

References

Aboubacar, A., and B.R. Hamaker. 1999. Physicochemical properties of flours that relate to sorghum couscous quality. Cereal Chem. 76:308–313. doi:10.1094/CCHEM.1999.76.2.308

Akingbala, J.O., and L.W. Rooney. 1987. Paste properties of sorghum flour and starches. J. Food Proc. Pres. 11:13–24.

Aldrich, P.R., and J. Doebley. 1992. Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theor. Appl. Genet. 85:293–302.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Anglani, C. 1998. Sorghum endosperm texture - A review. Plant Foods Hum. Nutr. 52(1):67–76. doi:10.1023/A:1007955102401

Bean, S.R., O.K. Chung, M.R. Tuinstra, J.F. Pedersen, and J. Erpelding. 2006. Evaluation of the single kernel characterization system (SKCS) for measurement of sorghum grain attributes. Cereal Chem. 83:108–113. doi:10.1094/CC-83-0108

Beta, T., H. Corke, L.W. Rooney, and J.R.N. Taylor. 2001. Starch properties as affected by sorghum grain chemistry. J. Sci. Food Agric. 81:245–251. doi:10.1002/1097-0010(20010115)81:2<245::AID-JSFA805>3.0.CO;2-S

Bettge, A.D., M.J. Giroux, and C.F. Morris. 2000. Susceptibility of waxy starch granules to mechanical damage. Cereal Chem. 77:750–753. doi:10.1094/CCHEM.2000.77.6.750

Bordes, J., C. Ravel, J. Le Gouis, A. Lapierre, G. Charmet, and F. Balfourier. 2011. Use of a global wheat core collection for association analysis of flour and dough quality traits. J. Cereal Sci. 54:137–147. doi:10.1016/j.jcs.2011.03.004

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Brown, P.J., S. Myles, and S. Kresovich. 2011. Genetic support for phenotype-based racial classification in sorghum. Crop Sci. 51:224–230. doi:10.2135/cropsci2010.03.0179

Bueso, F.J., R.D. Waniska, W.L. Rooney, and F.P. Bejosano. 2000. Activity of antifungal proteins against mold in sorghum caryopses in the field. J. Agric. Food Chem. 48:810–816. doi:10.1021/jf9909712

Cagampang, G.B., and A.W. Kirleis. 1984. Relationship of sorghum grain hardness to selected physical and chemical measurements of grain quality. Cereal Chem. 61:100–105.

Casa, A.M., G. Pressoir, P.J. Brown, S.E. Mitchell, W.L. Rooney, M.R. Tuinstra, C.D. Franks, and S. Kresovich. 2008. Community resources and strategies for association mapping in sorghum. Crop Sci. 48:30–40. doi:10.2135/cropsci2007.02.0080

Cavalli-Sforza, L.L., and A.W.F. Edwards. 1967. Phylogenetic analysis: Models and estimation procedures. Am. J. Hum. Genet. 19:233–257.

Chen, F., F.Y. Zhang, X.C. Xia, Z.D. Dong, and D.Q. Cui. 2012. Distribution of puroindoline alleles in bread wheat cultivars of the Yellow and Huai valley of China and discovery of a novel puroindoline a allele without PINA protein. Mol. Breed. 29:371–378. doi:10.1007/s11032-011-9553-2

Cook, J.P., M.D. McMullen, J.B. Holland, F. Tian, P. Bradbury, J. Ross-Ibarra, E.S. Buckler, and S.A. Flint-Garcia. 2012. Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol. 158:824–834. doi:10.1104/pp.111.185033

de Alencar Figueiredo, L., B. Sine, J. Chantereau, C. Mestres, G. Fliedel, J.F. Rami, J.C. Glaszmann, M. Deu, and B. Courtois. 2010. Variability of grain quality in sorghum: Association with polymorphism in *Sh2, Bt2, SssI, Ae1, Wx* and *O2*. Theor. Appl. Genet. 121:1171–1185. doi:10.1007/s00122-010-1380-z

De Wet, J.M.J. 1978. Systematics and evolution of Sorghum sect. Sorghum (Gramineae). Am. J. Bot. 65:477–484. doi:10.2307/2442706

Douglas, J.H., T.W. Sullivan, P.L. Bond, and F.J. Struwe. 1990. Nutrient composition and metabolizable energy values of selected grain sorghum varieties and yellow corn. Poult. Sci. 69:1147–1155. doi:10.3382/ps.0691147

Folkertsma, R.T., H.F.W. Rattunde, S. Chandra, G.S. Raju, and C.T. Hash. 2005. The pattern of genetic diversity of Guinea-race *Sorghum bicolor* (L.) Moench landraces as revealed with SSR markers. Theor. Appl. Genet. 111:399–409. doi:10.1007/s00122-005-1949-0

Fujita, N., R. Satoh, A. Hayashi, M. Kodama, R. Itoh, S. Aihara, and Y. Nakamura. 2011. Starch biosynthesis in rice endosperm requires the presence of either starch synthase I or IIIa. J. Exp. Bot. 62:4819–4831. doi:10.1093/jxb/err125

Goodstein, D.M., S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, and D.S. Rokhsar. 2012. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. 40:D1178–D1186.

Guzman, C., L. Caballero, M.A. Martin, and J.B. Alvarez. 2012. Molecular characterization and diversity of the Pina and Pinb genes in cultivated and wild diploid wheat. Mol. Breed. 30:69–78. doi:10.1007/s11032-011-9599-1

Hamblin, M.T., A.M. Casa, H. Sun, S.C. Murray, A.H. Paterson, C.F. Aquadro, and S. Kresovich. 2006. Challenges of detecting directional selection after a bottleneck: Lessons from *Sorghum bicolor*. Genetics 173:953–964. doi:10.1534/genetics.105.054312

Hamblin, M.T., M.G.S. Fernandez, A.M. Casa, S.E. Mitchell, A.H. Paterson, and S. Kresovich. 2005. Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass *Sorghum bicolor*. Genetics 171:1247–1256. doi:10.1534/genetics.105.041566

Hamblin, M.T., M.G.S. Fernandez, M.R. Tunistra, W.L. Rooney, and S. Kresovich. 2007. Sequence variation at candidate loci in the starch metabolism pathway in sorghum: Prospects for linkage disequilibrium mapping. Crop Sci. 47:S125–S134. doi:10.2135/cropsci2007.01.0054tpg

Hamblin, M.T., S.E. Mitchell, G.M. White, J. Gallego, R. Kukatla, R.A. Wing, A.H. Paterson, and S. Kresovich. 2004. Comparative

population genetics of the panicoid grasses: Sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. Genetics 167:471–483. doi:10.1534/genetics.167.1.471

Hardy, O.J., and X. Vekemans. 2002. SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2:618–620. doi:10.1046/j.1471-8286.2002.00305.x

Harlan, J.R., and J.M.J. De Wet. 1972. Simplified classification of cultivated sorghum. Crop Sci. 12:172. doi:10.2135/cropsci1972.0011183X001200020005x

Jambunathan, R., M.S. Kherdekar, and J.W. Stenhouse. 1992. Sorghum grain hardness and its relationship to mold susceptibility and mold resistance. J. Agric. Food Chem. 40:1403–1408. doi:10.1021/jf00020a023

Jennings, P., and J. Cock. 1977. Centres of origin of crops and their productivity. Econ. Bot. 31:51–54. doi:10.1007/BF02860652

Lee, W.J., J.F. Pedersen, and D.R. Shelton. 2002. Relationship of sorghum kernel size to physiochemical, milling, pasting, and cooking properties. Food Res. Int. 35:643–649. doi:10.1016/S0963-9969(01)00167-3

Lin, Z., X. Li, L.M. Shannon, C.T. Yeh, M.L. Wang, G. Bai, Z. Peng, J. Li, H.N. Trick, T.E. Clemente, J. Doebley, P.S. Schnable, M.R. Tuinstra, T.T. Tesso, F. White, and J. Yu. 2012. Parallel domestication of the Shattering1 genes in cereals. Nat. Genet. 44:720–724. doi:10.1038/ng.2281

Liu, K., and S.V. Muse. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. doi:10.1093/bioinformatics/bti282

Loiselle, B.A., V.L. Sork, J.D. Nason, and C. Graham. 1995. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am. J. Bot. 82:1420–1425. doi:10.2307/2445869

Martin, C.R., R. Rousser, and D.L. Brabec. 1993. Development of a single-kernel wheat characterization system. Trans. ASAE 36:1399–1404.

Morris, C.F., G.A. Greenblatt, A.D. Bettge, and H.I. Malkawi. 1994. Isolation and characterization of multiple forms of friabilin. J. Cereal Sci. 20:167–174. doi:10.1006/jcrs.1994.1056

Murphy, S.P., C.R. Simmons, and H.W. Bass. 2010. Structure and expression of the maize (*Zea mays* L.) SUN-domain protein gene family: Evidence for the existence of two divergent classes of SUN proteins in plants. BMC Plant Biol. 10:269. doi:10.1186/1471-2229-10-269

Murray, S.C., W.L. Rooney, M.T. Hamblin, S.E. Mitchell, and S. Kresovich. 2009. Sweet sorghum genetic diversity and association mapping for brix and height. Plant Gen. 2:48–62. doi:10.3835/plantgenome2008.10.0011

Murthy, B.R., M.K. Upadhyay, and P.I. Manchanda. 1967. Classification and cataloguing of a world collection of genetic stocks of Pennisetum. In: Catalogue of the world collection of sorghum and Pennisetum. Indian J. Genet. Plant Breed. 27:313.

Pasquini, C. 2003. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. J. Braz. Chem. Soc. 14:198–219. doi:10.1590/S0103-50532003000200006

Patterson, N., A.L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS Genet. 2:e190. doi:10.1371/journal.pgen.0020190

Pedersen, J.F., C.R. Martin, F.C. Felker, and J.L. Steele. 1996. Application of the single kernel wheat characterization technology to sorghum grain. Cereal Chem. 73:421–423.

Perumal, R., P. Krishnaramanujam, M.A. Menz, S. Katile, J. Dahlberg, C.W. Magill, and W.L. Rooney. 2007. Genetic diversity among sorghum races and working groups based on AFLPs and SSRs. Crop Sci. 47:1375–1383. doi:10.2135/cropsci2006.08.0532

Preiss, J., and M.N. Sivak. 1998. Biochemistry, molecular biology and regulation of starch synthesis. Genet. Eng. 20:177–223.

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association. Nat. Genet. 38:904–909. doi:10.1038/ng1847

Pritchard, J.K., M. Stephens, N.A. Rosenberg, and P. Donnelly. 2000. Association mapping in structured populations. Am. J. Hum. Genet. 67:170–181. doi:10.1086/302959

Reif, J.C., M. Gowda, H.P. Maurer, C.F.H. Longin, V. Korzun, E. Ebmeyer, R. Bothe, C. Pietsch, and T. Wurschum. 2011. Association mapping for quality traits in soft winter wheat. Theor. Appl. Genet. 122:961–970. doi:10.1007/s00122-010-1502-7

Salas Fernandez, M., I. Kapran, S. Souley, M. Abdou, I. Maiga, C. Acharya, M. Hamblin, and S. Kresovich. 2009. Collection and characterization of yellow endosperm sorghums from West Africa for biofortification. Genet. Resour. Crop Evol. 56:991–1000. doi:10.1007/s10722-009-9417-3

SAS Institute. 2000. SAS 9.1.3 help and documentation. SAS Institute Inc., Cary, NC.

Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78:629–644. doi:10.1086/502802

Shechter, Y., and J.M.J. De Wet. 1975. Comparative electrophoresis and isozyme analysis of seed proteins from cultivated races of sorghum. Am. J. Bot. 62:254–261. doi:10.2307/2441859

Shimbata, T., T. Inokuma, A. Sunohara, P. Vrinten, M. Saito, T. Takiya, and T. Nakamura. 2011. High levels of sugars and fructan in mature seed of sweet wheat lacking *GBSSI* and *SSIIa* enzymes. J. Agric. Food Chem. 59:4794–4800. doi:10.1021/jf200468c

Smith, C.W., and R.A. Frederiksen. 2000. Sorghum: Origin, history, technology and production John Wiley & Sons, New York, NY.

Snowden, J.D. 1936. The cultivated races of sorghum. Adlard and Sons, London, UK.

Stephens, J.C., F.R. Miller, and D.T. Rosenow. 1967. Conversion of alien sorghums to early combine genotypes. Crop Sci. 7:396. doi:10.2135/cropsci1967.0011183X000700040036x

Suhendro, E.L., C.F. Kunetz, C.M. McDonough, L.W. Rooney, and R.D. Waniska. 2000. Cooking characteristics and quality of noodles from food sorghum. Cereal Chem. 77:96–100. doi:10.1094/CCHEM.2000.77.2.96

Sun, G., C. Zhu, M.H. Kramer, S.S. Yang, W. Song, H.P. Piepho, and J. Yu. 2010. Variation explained in mixed-model association mapping. Heredity 105:333–340. doi:10.1038/hdy.2010.11

Taylor, J.R.N., T.J. Schober, and S.R. Bean. 2006. Novel food and non-food uses for sorghum and millets. J. Cereal Sci. 44:252–271. doi:10.1016/j.jcs.2006.06.009

Tian, Z., Q. Qian, Q. Liu, M. Yan, X. Liu, C. Yan, G. Liu, Z. Gao, S. Tang, D. Zeng, Y. Wang, J. Yu, M. Gu, and J. Li. 2009. Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. Proc. Natl. Acad. Sci. USA 106:21760–21765. doi:10.1073/pnas.0912396106

Wills, R.B.H., and M.R. Ali. 1983a. Effect of grain-size on degree of milling, color and cooking time of sorghum. J. Food Sci. 48:650–651. doi:10.1111/j.1365-2621.1983.tb10813.x

Wills, R.B.H., and M.R. Ali. 1983b. Effect of grain-size on dehulling of sorghum. Cereal Chem. 60:12–14.

Wilson, L.M., S.R. Whitt, A.M. Ibanez, T.R. Rocheford, M.M. Goodman, and E.S. Buckler. 2004. Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733. doi:10.1105/tpc.104.025700

Yan, H.B., X.X. Pan, H.W. Jiang, and G.J. Wu. 2009. Comparison of the starch synthesis genes between maize and rice: Copies, chromosome location and expression divergence. Theor. Appl. Genet. 119:815–825. doi:10.1007/s00122-009-1091-5

Yan, L., X. Peng, S. Yafang, Z. Gan, T. Thanwanit, and B. Jinsong. 2010. Development of new markers to genotype the functional SNPs of SSIIa, a gene responsible for gelatinization temperature of rice starch. J. Cereal Sci. 52:438–443. doi:10.1016/j.jcs.2010.07.008

Yu, J., M.T. Hamblin, and M.R. Tuinstra. 2011. Association genetics strategies and resources. In: A.H. Paterson, editor, Genetics and genomics of Saccharinae. Springer-Science, New York, NY.

Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208. doi:10.1038/ng1702

Zhu, C., X. Li, and J. Yu. 2011. Integrating rare-variant testing, function prediction, and gene network in composite resequencing-based genome-wide association studies (CRGWAS). G3. 1:233–243.

Zhu, C., and J. Yu. 2009. Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. Genetics 182:875–888. doi:10.1534/genetics.108.098863